Chapter_5_HW

Christophe

October 12, 2015

5.6 Working backwards, Part II.

A 90% cofidence interval for a population mean is (65, 77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

The sample mean is the midpoint at $\bar{x} = 71$

The margin of error is 6

The degrees of freedom is `n - 1` so for this instance it is 24 and the z score from the t-table is `1.711`. Now in order to calculate the standard deviation we take the margin of error divided by the applicable z score and multiplied by the square root of `n` so in this instance $\sigma = 17.534$

5.14 SAT scores.

SAT scores of students at an Ivy League college are distributed with a standard deviation of 250 points. Two statistics students, Raina and Luke, want to estimate the average SAT score of students at this college as part of a class project. They want their margin of error to be no more than 25 points.

(a) Raina wants to use a 90% confidence interval. How large a sample should she collect?

$\sigma = 250$

$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

Margin of error ≤ 25

Margin of error formula: $Z * SE$

$25 = 1.645 * \frac{250}{\sqrt{n}}$

$25 = \frac{411.25}{\sqrt{n}}$

$625 = \frac{1.6912656 \times 10^5}{n}$

$625n = 1.6912656 \times 10^5$

$n = 270.6025$

$n_{people} = 271$

(b) Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning.

Luke would need a larger sample because he wants to increase the confidence of having captured the true mean of the population. By increasing the confidence it would require more observations to ensure the true mean is captured.

(c) Calculate the minimum required sample size for Luke.

$\sigma = 250$

$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

Margin of error ≤ 25

Margin of error formula: $Z * SE$

$25 = 2.575 * \frac{250}{\sqrt{n}}$

$25 = \frac{643.75}{\sqrt{n}}$

$625 = \frac{4.1441406 \times 10^5}{n}$

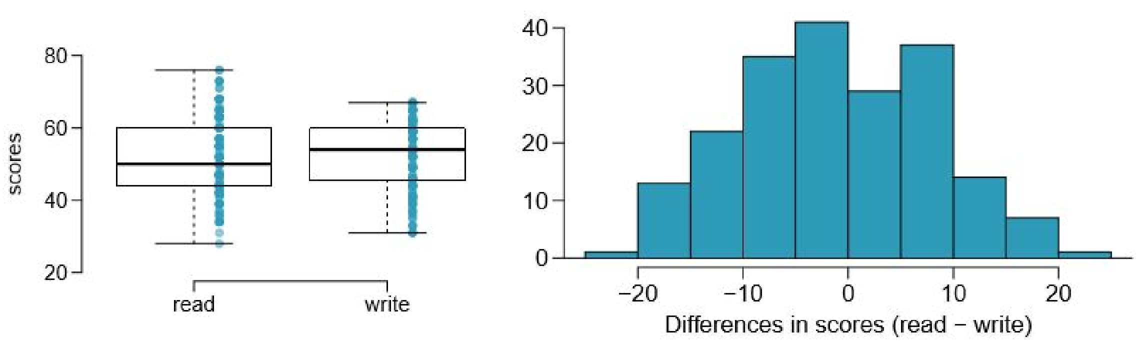
$625n = 4.1441406 \times 10^5$

$n = 663.0625$

$n_{people_Luke} = 664$

5.20 High School and Beyond, Part I.

The National Center of Education Statistics conducted a survey of high school seniors, collecting test data on reading, writing, and several other subjects. Here we examine a simple random sample of 200 students from this survey. Side-by-side box plots of reading and writing scores as well as a histogram of the differences in scores are shown below.



(a) Is there a clear difference in the average reading and writing scores?

There appears to be some differences, while the write portion has higher mean the range of scores is much more narrow than the reading portion.

(b) Are the reading and writing scores of each student independent of each other?

I would not think that they are independent because a student that can read would also have the same quality of writing ability. Essentially, having information about one variable would give me some information about the other and therefore not independent.

(c) Create hypotheses appropriate for the following research question: is there an evident difference in the average scores of students in the reading and writing exam?

$$H_o : \mu_{reading} = \mu_{writing}$$
$$H_A : \mu_{reading} \neq \mu_{writing}$$

(d) Check the conditions required to complete this test.

We are able check the conditions that this is less than 10% of the population because 200 is less than 10% of all high school students. We are also able to check that the distribution appears to be near normal. The sample is greater than 30 so we can say that this passes the conditions required to complete the test.

(e) The average observed difference in scores is $\bar{x}_{read-write} = -0.545$, and the standard deviation of the differences is 8.887 points. Do these data provide convincing evidence of a difference between the average scores on the two exams?

$$T - score = \frac{-0.545 - 0}{8.887} = -0.0613255$$

T score for df = 199 is 1.97.

Since -0.0613255 < 1.97 we can conlude there is a difference in observed scores. Sufficient evidence exists that the two exams have a difference.

(f) What type of error might we have made? Explain what the error means in the context of the application.

We may have made a type 1 error where we rejected the null hypothesis when it was in fact true because the expected mean of reading and the expected mean of writing do not equal each other.

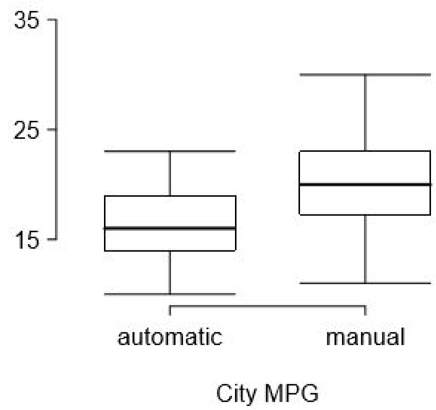
(g) Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the reading and writing scores to include 0? Explain your reasoning.

I would expect the confidence interval of the average differences to include 0. We can see from the differences in the histogram that the differences are near normally distributed around 0 so a confidence interval should include 0.

5.32 Fuel efficiency of manual and automatic cars, Part I.

Each year the US Environmental Protection Agency (EPA) releases fuel economy data on cars manufactured in that year. Below are summary statistics on fuel efficiency (in miles/gallon) from random samples of cars with manual and automatic transmissions manufactured in 2012. Do these data provide strong evidence of a difference between the average fuel efficiency of cars with manual and automatic transmissions in terms of their average city mileage? Assume that conditions for inference are satisfied.

	City MPG	
	Automatic	Manual
Mean	16.12	19.85
SD	3.58	4.51
n	26	26



$$\bar{x}_{automatic} - \bar{x}_{manual} = -3.73$$

$$SE_{\bar{x}_{automatic}-\bar{x}_{manual}} = \sqrt{\frac{\sigma^2_{automatic}}{n_{automatic}} + \frac{\sigma^2_{manual}}{n_{manual}}}$$

$$SE_{\bar{x}_{automatic}-\bar{x}_{manual}} = \sqrt{\frac{16.12^2}{26} + \frac{19.85^2}{26}}$$

$$SE_{\bar{x}_{automatic}-\bar{x}_{manual}} = \sqrt{\frac{259.8544}{26} + \frac{394.0225}{26}}$$

$$SE_{\bar{x}_{automatic}-\bar{x}_{manual}} = \sqrt{9.9944 + 15.1547115}$$

$$SE_{\bar{x}_{automatic}-\bar{x}_{manual}} = 5.014889$$

df = 25

$$T - score = \frac{-3.73 - 0}{5.014889} = -0.7437852$$

t value from t table = 2.06, assuming a two tail test and 95% confidence interval.

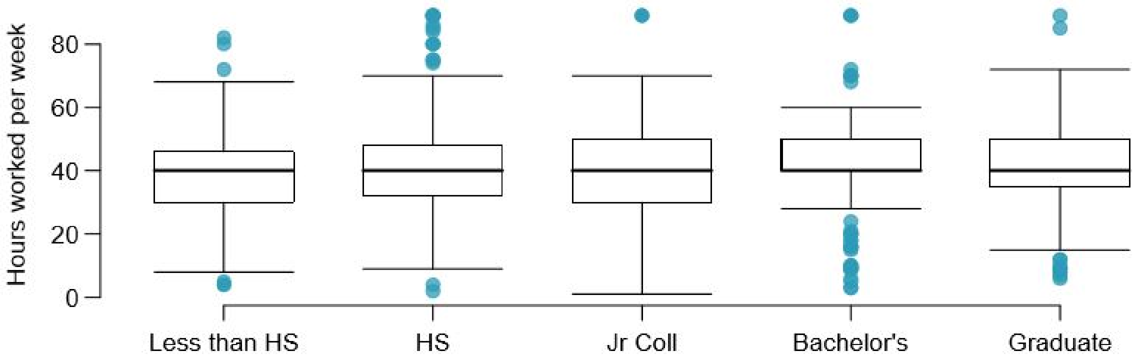
$0.7437852 < 2.06$

Therefore, because our p is less than our t-score at 95% significance we reject H_o and can conclude that there is strong evidence that there is a difference in gas mileage between different transmission types.

5.48 Work hours and education.

The General Social Survey collects data on demographics, education, and work, among many other characteristics of US residents. Using ANOVA, we can consider educational attainment levels for all 1,172 respondents at once. Below are the distributions of hours worked by educational attainment and relevant summary statistics that will be helpful in carrying out this analysis

	Educational attainment					
	Less than HS	HS	Jr Coll	Bachelor's	Graduate	Total
Mean	38.67	39.6	41.39	42.55	40.85	40.45
SD	15.81	14.97	18.1	13.62	15.51	15.17
n	121	546	97	253	155	1,172



(a) Write hypotheses for evaluating whether the average number of hours worked varies across the five groups.

$$H_o : \mu_{lessthanHS} = \mu_{HS} = \mu jrcoll = \mu bachelors = \mu graduate$$
$$H_A : \mu_{lessthanHS} \neq \mu_{HS} \neq \mu jrcoll \neq \mu bachelors \neq \mu graduate$$

(b) Check conditions and describe any assumptions you must make to proceed with the test.

We know that 1,172 is less than 10% of the population of the US. The sample is much greater than 30 so some skew is acceptable, so assuming there are no other causes for concern we can say this passes the conditions required to complete the test.

(c) Below is part of the output associated with this test. Fill in the empty cells.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree			501.54		0.0682
Residuals		267,382			
Total					

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
degree	4	2,006.16	501.54	2.196	0.0682
Residuals	1171	267,382	228.34		
Total	1175				

(d) What is the conclusion of the test?

Pr(>F) is `0.0682` or 6.82% if we choose a level σ of 5% we can not reject H_o, but if we choose a level σ of 7% or greater we can reject H_o.

Therefore, assuming a σ of 5% we fail to reject the null hypothesis and the mean number of worked across educational attainment appears similar.